

Comparison of neighbourhoods in New York, Toronto and San Francisco

David Golob

June 27, 2020

Introduction

- ▶ **Purpose:** Compare cities based on neighbourhood features.
- ▶ **Idea:** The idea is that we obtain the data about venues in each of cities' neighbourhoods. Based on this data we can cluster neighbourhoods in each city, then we can compare cluster centroids between each city and calculate similarity measure.

Data

- ▶ Postal codes, neighbourhood names, latitude and longitude obtained from the following urls:
 - ▶ <https://www.geonames.org/postal-codes/US/CA/075/san-francisco.html>
 - ▶ <http://www.healthysf.org/bdi/outcomes/zipmap.htm>
 - ▶ https://coc1.us/new_york_dataset
 - ▶ https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
 - ▶ https://coc1.us/Geospatial_data
- ▶ The venue data was obtained from **foursquare** api.

Methodology

- ▶ The data we use for each city are the neighbourhoods in this city and the percentage of each venue type in each neighbourhood.
- ▶ We apply k-means clustering algorithm to this data, where we take $k = 5$.
- ▶ Once we have clusters for each city we can determine similarity between clusters in different cities.
- ▶ The algorithm for determination of most similar clusters is the following:
 1. We pick the primary city and secondary city.
 2. For each cluster center for primary city we find the secondary city's cluster center that is the nearest (where we use euclidean distance).
 3. We report the primary city's cluster centers, their nearest secondary city's cluster centers and the distance between the two.

Calculation of similarity between cities

- ▶ The distance between two cities is the sum of distances obtained in the previous algorithm where we have to take into account that both cities need to be primary and secondary city once.

Results

In this section we present similarities between cities.

Toronto and San Francisco

1. The distance between Toronto and San Francisco :
5.3346

	San Francisco Cluster	Toronto Cluster	Distance
0	0	0	0.14579
1	1	0	0.456625
2	2	0	0.415387
3	3	0	0.240626
4	4	0	0.319868

Figure: San Francisco vs. Toronto

	Toronto Cluster	San Francisco Cluster	Distance
0	0	0	0.14579
1	1	0	0.991026
2	2	0	0.996593
3	3	0	0.61734
4	4	0	1.0056

Figure: Toronto vs. San Francisco

Toronto and New York

1. The distance between Toronto and New York : 5.8688

	Toronto Cluster	New York Cluster	Distance
0	0	0	0.16114
1	1	0	1.00787
2	2	1	0.995878
3	3	0	0.600508
4	4	1	1.00608

Figure: Toronto vs. New York

	New York Cluster	Toronto Cluster	Distance
0	0	0	0.16114
1	1	0	0.166457
2	2	0	0.994531
3	3	0	0.582137
4	4	0	0.193151

Figure: New York vs. Toronto

New York and San Francisco

1. **The distance between New York and San Francisco :**
3.6641

	New York Cluster	San Francisco Cluster	Distance
0	0	0	0.155735
1	1	0	0.148697
2	2	0	1.00178
3	3	0	0.579033
4	4	0	0.187758

Figure: New York vs San Francisco

	San Francisco Cluster	New York Cluster	Distance
0	0	1	0.148697
1	1	0	0.453666
2	2	0	0.426777
3	3	0	0.237369
4	4	0	0.324605

Figure: San Francisco vs. New York

Discussion

- ▶ Other possibilities of calculation of similarities between cities
- ▶ Maybe try with different values of parameter k in k-means clustering.

Conclusion

- ▶ New York is more similar to San Francisco than to Toronto